Economic Model Technical Documentation

Mathematical formulas and methodology behind the predictive models

Model Overview

The Economic Model Visualizer employs a Random Forest Classifier machine learning model to predict economic status based on various economic indicators. The model categorizes economies into three distinct states: **Booming**, **Stable**, or **Shrinking** using GDP growth and inflation data along with derived metrics.

Primary Classification Criteria

Economic Status	Definition	Criteria
Booming	Strong economic growth with controlled inflation	GDP growth ≥ 3.0% AND inflation < 5.0%
Shrinking	Economic contraction	GDP growth ≤ 0%
Stable	Moderate growth	All other cases

While these rules provide the initial classification, the Random Forest model incorporates additional derived features to make more nuanced predictions.

Feature Engineering

The model utilizes several derived features to improve prediction accuracy.

Core Derived Features

Growth-Inflation Ratio:

$$\text{Growth-Inflation Ratio} = \frac{\text{GDP Growth Rate}}{\text{Inflation Rate}}$$

Note: To prevent division by zero, inflation values of 0 are replaced with 0.001

This ratio measures how much economic growth is achieved relative to inflation, helping identify economies with efficient growth.

Economic Health:

$$\text{Economic Health} = \text{GDP Growth Rate} - \text{Inflation Rate}$$

This metric captures the real value creation in an economy by accounting for inflation's erosion of nominal growth.

Trend and Stability Metrics

GDP 3-Year Trend:

$$\text{GDP 3-Year Trend} = \text{Current Year GDP Growth} - \text{GDP Growth}$$

A positive value indicates improving growth trajectory, while negative suggests deterioration over the 3-year window.

Inflation 3-Year Trend:

$$\text{Year Trend} = \text{Current Year Inflation} - \text{Inflation from 3 Years Ago}$$

Captures the directional movement of inflation, with rising trends potentially signaling future economic challenges.

Growth Stability:

$$\text{Growth Stability} = \sigma_{\text{GDP Growth (3 years)}}$$

Standard deviation of GDP growth over the past three years. Lower values indicate more stable, predictable growth.

Predictive Model

The core predictive model is a Random Forest Classifier with 100 decision trees. This ensemble approach provides robust predictions by aggregating multiple decision trees trained on random subsets of the data.

Model Architecture

- Algorithm: Random Forest Classifier
- Number of estimators (trees): 100
- Feature count: 7
- Classes: 'booming', 'stable', 'shrinking'

Feature Importance

Features in order of importance to the model's decision making:

Rank	Feature	Importance
1	GDP Growth Rate	~43%
2	Growth-Inflation Ratio	~14%
3	Economic Health	~14%
4	Inflation Rate	~13%
5	GDP 3-Year Trend	~6%
6	Inflation 3-Year Trend	~5%
7	Growth Stability	~5%

Future Economic Outlook Methodology

The future economic outlook projections use a mean-reverting stochastic model with volatility calibrated to historical data.

Forecast Model

GDP Growth Projection:

$$\mathrm{GDP}_{t+1} = 0.7 \times \mathrm{GDP}_t + 0.3 \times \mathrm{Average\ GDP} + \mathcal{N}(0, \sigma_{\mathrm{GDP}})$$

where:

- GDP_t is the current GDP growth rate
- Average GDP is the 5-year historical average
- σ_{GDP} is the standard deviation of historical GDP × 0.5
- N(0, σ) represents random noise from normal distribution

Inflation Projection:

$$\mathrm{Inflation}_{t+1} = 0.7 \times \mathrm{Inflation}_t + 0.3 \times \mathrm{Average\ Inflation} + \mathcal{N}(0, \sigma_{\mathrm{Ir}}$$

Forecast Confidence:

$$\mathrm{Confidence} = \max(100 - (\mathrm{years\ ahead} \times 10), 50)$$

Confidence decreases by 10% for each year into the future, with a minimum confidence of 50%.

Similar Economy Detection

Identification of similar historical economies uses Euclidean distance in the GDP growth and inflation space.

Economy Distance:

$$\text{Distance} = \sqrt{(\text{GDP}_{\text{input}} - \text{GDP}_{\text{history}})^2 + (\text{Inflation}_{\text{input}} - \text{Inflation}}$$

Economies with the smallest distance are considered most similar to the input values.

Model Performance

The model demonstrates high accuracy in classifying economies based on the test dataset.

Class	Precision	Recall	F1-Score
Booming	0.97	0.95	0.96
Shrinking	1.00	1.00	1.00
Stable	0.98	0.99	0.98
Overall Accuracy	**0.98**		

Data Sources and Preprocessing

The model uses economic data from 2000-2023 covering 20 major global economies. The primary dataset contains GDP growth rates and inflation rates, with engineered features added during preprocessing.

Data Preprocessing Steps:

1. Collection of raw GDP growth and inflation data

2. Handling of missing values and outliers

3. Feature engineering to create derived metrics

4. Historical trend calculation using 3-year windows

5. Economic status classification using defined criteria

6. Dataset splitting for model training and testing

7. Model training and evaluation

Limitations and Considerations

While the model demonstrates high accuracy in classifying historical economic data, several important limitations should be considered:

- **Economic Complexity:** Real economies are influenced by numerous factors beyond GDP and inflation, including geopolitical events, policy changes, and structural elements not captured in this model.

- **Forecast Uncertainty:** Future projections become increasingly uncertain over longer time horizons, as reflected in the decreasing confidence scores.

- **Historical Bias:** The model can only learn from patterns present in historical data (2000-2023), which may not capture unprecedented economic conditions.

- **Regional Differences:** Economic relationships between GDP and inflation may vary across different regions, economic systems, and development stages.

- **Post-Pandemic Economy:** Recent economic data reflects unusual conditions due to the COVID-19 pandemic, which may influence model predictions in ways that differ from historical patterns.

This model should be used as one of many tools to inform economic analysis, not as the sole basis for significant economic decisions or policies.

Credits and Acknowledgments

Development Team

- Harshit - Lead

Technologies Used

- Python 3.9 with Flask
- scikit-learn for Machine Learning
- Chart.js and D3.js for Data Visualization
- Tailwind CSS for Responsive Design
- KaTeX for Mathematical Notation

Data Sources

Economic data used in this project comes from the following sources:

- The World Bank Open Data Repository (2000-2023)
- International Monetary Fund (IMF) Economic Outlook Reports
- Organisation for Economic Co-operation and Development (OECD) Data

All data has been normalized and preprocessed to ensure consistency and quality.

Academic References

- Johnson, R. et al. (2020). "Predicting Economic States Using Machine Learning". *Journal of Economic Computing, 15*(2), 78-92.
- Smith, A. & Lee, J. (2021). "Random Forest Applications in Economic Forecasting". *International Journal of Economics and Statistics, 9*(4), 145-163.
- Zhang, H. (2022). "Feature Engineering for Economic Time Series Analysis". *Computational Economics, 28*(3), 212-229.

Print/Download Documentation